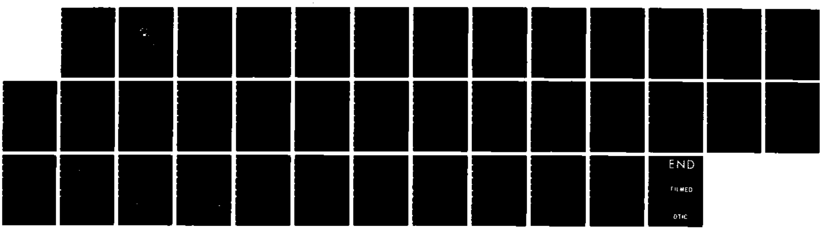
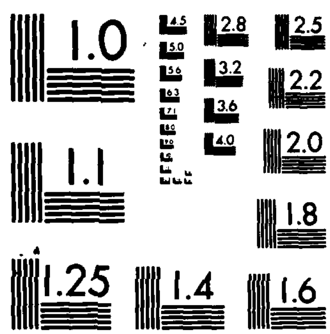
AD-A159 336 LAPLACIAN SMOOTHING SPLINES WITH GENERALIZED CROSS VALIDATION FOR OBJECTI.. (U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA R FRANKE AUG 85 NPS-53-85-0008 1/1

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A159 336

NPS-53-85-0008

NAVAL POSTGRADUATE SCHOOL

Monterey, California

DTIC
ELECTE
SEP 24 1985
S E D

LAPLACIAN SMOOTHING SPLINES WITH
GENERALIZED CROSS VALIDATION FOR
OBJECTIVE ANALYSIS OF METEOROLOGICAL DATA

by

Richard Franke

August 1985

Technical Report For Period

October 1984 - March 1985

Approved for public release; distribution unlimited

Prepared for:
Naval Environmental Prediction Research Facility
Monterey, CA 93943

85 09 23 044

NAVAL POSTGRADUATE SCHOOL
MONTEREY CALIFORNIA 93943

R. H. Shumaker
Rear Admiral, U. S. Navy
Superintendent

D. A. Schrady
Provost

This work was funded by the Naval Environmental Prediction Research Facility, Monterey, CA under Program Element 61153N, Project (none), "Interpolation of Scattered Meteorological Data".

Reproduction of all or part of this report is authorized.

Richard Franke
Professor of Mathematics

Reviewed by:

Gordon E. Latta, Chairman
Department of Mathematics

Released by:

Kneale T. Marshall
Dean of Information and Policy
Sciences

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NPS-53-85-0008	2. GOVT ACCESSION NO. AD-A159	3. RECIPIENT'S CATALOG NUMBER 336
4. TITLE (and Subtitle) Laplacian Smoothing Splines with Generalized Cross Validation for Objective Analysis of Meteorological Data		5. TYPE OF REPORT & PERIOD COVERED Technical Report Interim, FY 1985 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard Franke		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element 61153N
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Environmental Prediction Research Facility Monterey, CA 93943		12. REPORT DATE August 1985 13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Air Systems Command Washington, DC 20361		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Objective analysis
Laplacian smoothing splines
Optimum interpolation
Generalized cross validation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The use of Laplacian smoothing splines (LSS) with generalized cross validation (GCV) to choose the smoothing parameter for the objective analysis problem is investigated. Simulated 500 mb pressure height fields are approximated from first-guess data with spatially correlated errors and observed values having independent errors. It is found that GCV does not allow LSS to adapt to variations in individual realizations, and that specification of a single suitable smoothing parameter value for all

BLOCK 20 - ABSTRACT

realizations leads to smaller rms error overall. While the tests were performed in the context of data from a meteorology problem, it is expected the results carry over to data from other sources. A comparison shows that significantly better approximations can be obtained using LSS applied in a unified manner to both first-guess and observed values rather that in a correction to first-guess scheme (as in Optimum Interpolation) when the first-guess error has low spatial correlation.

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC COPY INSPECTED

1. Introduction

In numerical weather prediction, objective analysis is the process of combining information obtained from observations of meteorological variables with that from the numerical prediction process. The resulting "analyzed" values are used to prepare weather maps, as well as to initialize the variables for the next weather prediction cycle. The problem is inherently a multivariate one since the variables are not independent, e.g., pressure heights are related to winds. The predicted values are on a regular grid, and have errors which are spatially correlated. The observed values are measured imperfectly, and occur at irregularly spaced (scattered) points (both in space and time). The errors in the observations sometimes occur independently, with zero mean, and in other cases, such as satellite observations, are biased with correlated errors.

The traditional approach to the problem is a two step process. The predicted values are treated as a first-guess and interpolated from the grid to the observation points. The difference between the first-guess values interpolated to the observation points and the observed values, called the first-guess error, is then interpolated back to the grid points as a correction to the first-guess values. The interpolation from grid-to-observation points is the "easy" process, and has not received much attention in the literature. The procedure generally used is multilinear interpolation (e.g., Bergman, 1979, or Lorenc, 1981), although recent investigations by the author (Franke, 1985) have demonstrated that appreciable error may occur in this

step. The interpolation from observation-to-grid points is the "hard" problem and has received widespread attention. Historically the favored scheme has been a weighted average scheme, originally introduced by Cressman (1959), with a variation due to Barnes (1973). Currently the method of choice is a statistical scheme known in the meteorological literature as Optimum Interpolation (OI), and in other disciplines by other names (e.g., Kriging in the mining and geology literature).

The interpolation process known as OI has its roots in the work of Weiner and Kolmogorov, and was introduced to the meteorological literature by Gandin (1963). The theory of the process depends on it being applied to a random function with known spatial statistics. In particular it is assumed that the spatial covariance structure of the class of functions to which it is applied is known. In addition it is necessary to know the error statistics of the observation devices. If this is the case, then the process yields the best answer possible in the sense that the variance of the error is minimized over all functions in the class. For meteorological purposes, this means the covariance structure of an ensemble of realizations must be known, and then the mean squared error over the entire ensemble is minimized. Using standard least squares methods, the variance of the expected error is easily computed, and much emphasis has been put on this as an advantage of the method.

There have been numerous papers about the multivariate application of OI to the objective analysis problem. These are of an applications nature, and it is difficult to separate the behavior of such schemes from that of the other involved

processes. In studies of objective analysis using simulated data to attempt to learn something about the properties of the scheme, many simplifications are required. This study is no different. The univariate (only one meteorological variable is treated, in this case the 500 mb pressure height surface) application of OI and other schemes is investigated. Because the generation of simulated data with specified spatial correlation properties requires the factorization of the correlation matrix for the first guess error at the grid points, it is necessary to work with a relatively small grid. Further, the problem of non-synoptic observation of variables is not treated, rather all observations are assumed made at the same time, the time at which the particular realization occurs. Within the prescribed limitations, the procedure used is valid and yields information about the objective analysis process which should prove to be useful in practice.

A somewhat different way of looking at the problem was proposed by Wahba and Wendelberger (1980). See also Wendelberger (1981). In their work, no first guess was necessary or assumed; all data was considered to be observation values. Thus the underlying field to be approximated was treated directly, rather than making a correction to the first-guess field. The overall process involved the use of Laplacian smoothing splines and generalized cross validation to determine a suitable value for the smoothing parameter. If a first guess is available, with known correlated errors, then ignoring this information is probably unwise. The first-guess can be used in the traditional manner, with the Laplacian smoothing splines applied to the

first-guess error. It is also possible to apply the Laplacian smoothing splines to all of the data. Thus, part of the investigation reported here involved the use of Laplacian smoothing splines and generalized cross validation for the smoothing parameter in a scheme that approximates the underlying field directly, but that also makes use of all available data in a way that accounts for the correlation of the errors. The program used was a modified version of the program MSSP, available from the Madison Academic Computing Center, University of Wisconsin.

Section 2 gives an outline of the goals of this study, background information about the methods of objective analysis considered, and aspects of the schemes investigated. The results of the study are given and discussed in Section 3. Finally, the implications of the results and conclusions about approaches to objective analysis, and suggestions for further study are given in Section 4.

2. Goals of the study

This study had two principal goals: (1) To investigate the efficacy of generalized cross validation (GCV) in determining the smoothing parameter used in Laplacian smoothing splines (LSS), and (2) To test the possibility of treating first-guess values and observed values in a unified method with LSS. The smoothing parameter value must be given in order to use LSS, and Wahba and Wendelberger (1980) have indicated that GCV might be a good way to choose the value. In this study I performed simulations to determine if GCV could adapt properly to particular realizations in an ensemble with specified error statistics.

The advantage of a unified scheme for both first-guess and observed values is that it potentially makes it possible to obtain better analyses where the observations are sparse compared to the grid or correlation distances. The LSS method used in this investigation was the scheme proposed by Wahba and Wendelberger (1980), which is described more fully in Wendelberger (1981, 1982). The general framework of this study follows that of a previous investigation (Franke, 1985).

A brief description of the setting in which the numerical experiments were performed follows. An underlying function to be approximated was chosen. The simulated pressure height field described by Koehler (1979) was used, at the 500 mb level, with random values for two parameters, θ_o (chosen uniformly distributed on $[-112.5^o,-82.5^o]$), and $\Delta\theta$ (chosen uniformly distributed on $[-15^o,15^o]$). One possible realization of the field is shown in Figure 4. The underlying field was then evaluated on a rectangular grid. Normally distributed first-guess errors with specified spatial covariance were generated and added to the field values to obtain the first-guess values. Then, the underlying field was evaluated at a set of observation points, and normally distributed independent observation errors with specified variance were added to these values to obtain observation values. An objective analysis scheme was then applied using the first-guess and observation values to obtain estimates of the underlying field at the grid points; these are called the analyzed values of the field. The errors in the analyzed values were then computed. After repeating the process for many realizations, estimates of the root-mean-square error was obtained.

In order to avoid edge effects, rms errors for the first-guess and analyzed values were tabulated only over the interior grid points. In a previous study (Franke, 1985), this process was used to obtain simulated results using various objective analysis schemes, under various assumptions about parameters in statistical schemes and other methods. In the current study this process is the starting point for investigations indicated above.

The approach taken for OI is to view the approximation as a linear combination of the spatial covariance functions for the observation points,

$$F(P) = \sum_{k=1}^{N_O} a_k C(P,P_k)$$

Here $C(P,Q)$ is the stationary, isotropic covariance function for the first-guess error, $F(P)$ is the approximating function, the observation points are P_k, with first guess values F_k, $k=1,\ldots,N_O$, and the a_k satisfy the system of equations

$$\sum_{k=1}^{N_O} a_k \left(C\,(P_i,P_k) + \delta_{ik} r_k^{\,2}\right) = F_i,\ i=1,\ldots,\ N_O.$$

The r_k denote the standard deviation of the observation errors at P_k.

One of the practical difficulties of the method is the specification of a suitable covariance structure. Not only is this important from the standpoint of modeling the process properly, but also from the standpoint of obtaining meaningful estimates of the mean squared error. In fact, these estimates

hold only when the covariance structure is known. When a particular structure is assumed, with parameter values being estimated from a time history or otherwise specified inexactly, these estimates may differ substantially from the actual values (Franke, 1985). As a matter of terminology, it is noted that when the process is applied using empirically derived, or assumed covariance functions, the scheme is called "statistical interpolation" in the meteorology literature. It is easily observed that the accuracy of the scheme is closely related to a somewhat nebulous quantity which I will refer to as the "correlation distance". This quantity indicates something about the distance at which the spatial correlation in the first-guess values drops below a certain level. If the distance from observation points to the analysis point (a grid point, in this case) is greater than the correlation distance, then the scheme cannot perform well, and in fact may only improve the value slightly. Thus the performance of the method is strongly dependent on the first-guess errors being correlated, the the higher the correlation, the better.

The scheme proposed by Wahba and Wendelberger (1980) is based on the use of LSS. These functions were first introduced as interpolation functions by Harder and Desmarais (1972), and were later developed more fully by Duchon (1976, 1977) and Meinguet (1979, 1979a). The generalization to smoothing and their application to the objective analysis problem was by Wahba and Wendelberger. The functions obtain their name, and are characterized by minimization of a functional related to the iterated Laplacian, Δ^m;

$$\lambda \sum_{i+j=m} \binom{m}{i} \iint_{R^2} \left(\frac{\partial^m H}{\partial\theta^i \partial\phi^j} \right)^2 dA + N_0^{-1} (\Delta H)^t \Sigma^{-1} (\Delta H).$$

Here λ is a smoothing parameter, and the order of the LSS m (>1) determines the smoothness of the function in terms of the number of continuous derivatives it must have. The N_o vector ΔH is differences between the approximation values and the data values, and Σ is the covariance matrix between the errors in the data taken over an ensemble of realizations. In the context of the objective analysis problem being considered, the solution of the problem can be shown to be a function of the form

$$H(P) = \sum_{k=1}^{No} A_k \, B(P,P_k) + \sum_{i+j<m} b_{ij} \, \theta^i \phi^j ,$$

where the independent variables are taken to be longitude, θ, and latitude, ϕ, and the data points are (P_k,H_k) with $P_k = (\theta_k,\phi_k)$, $k=1,\ldots,N_o$. The basis functions for the approximation depend on the number of independent variables. For the case of two independent variables the basis functions can be taken to be $B(P,Q) = ||P-Q||^{2m-2}\log||P-Q||$, where $||P-Q||$ is the distance in degrees between points P and Q. The coefficients A_k, and those of the polynomial $\sum_{i+j<m} b_{i,j}\theta^i\phi^j$, satisfy the system of equations

$$\sum_{k=1}^{N_o} A_k \left((B\,(P_n,P_k) + \Sigma_{n,k}\right) + \sum_{i+j<m} b_{ij}\, \theta_n^i \,\phi_n^j, \quad n=1,\ldots,N_o$$

$$\sum_{k=1}^{No} A_k \,\theta_k^i \,\phi_k^j = 0 \;, \quad i+j<m$$

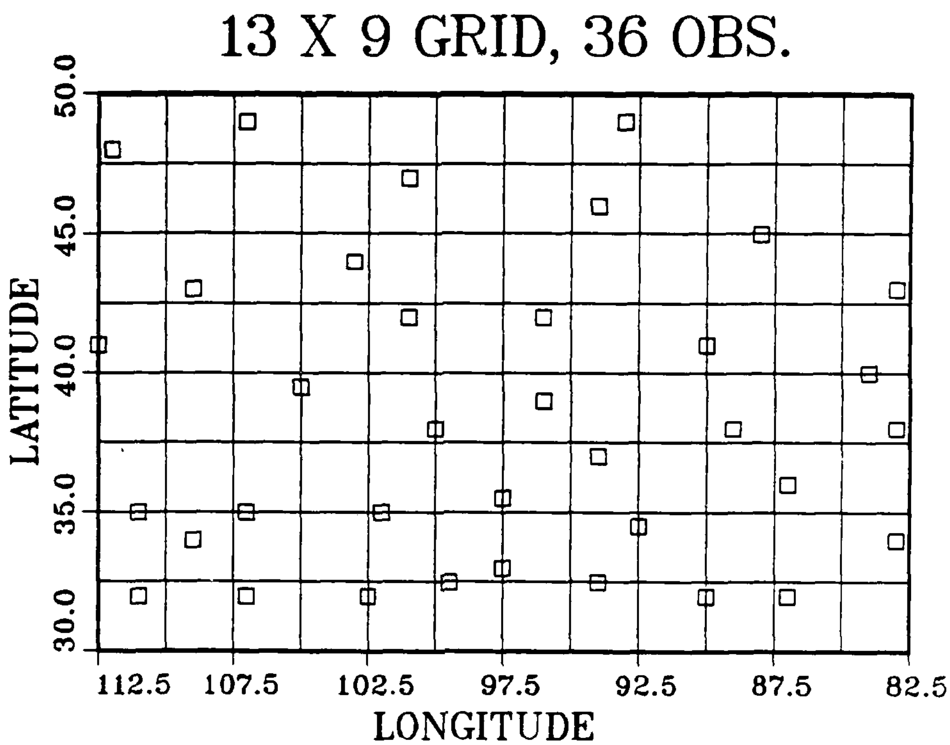


FIGURE 1

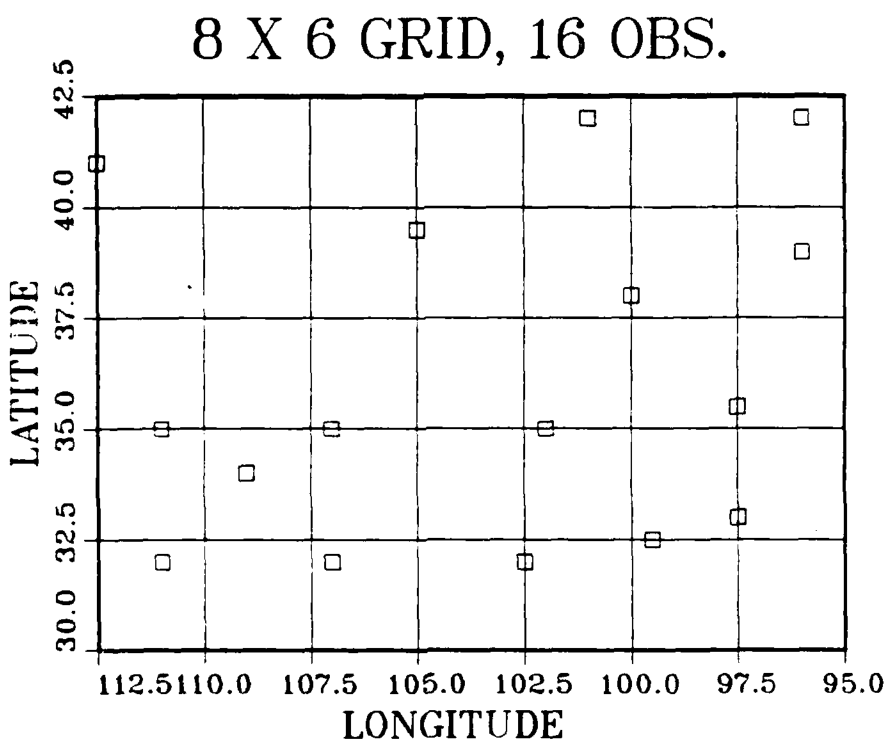


FIGURE 2

Grid	#Obs	c_d	GCV2	GCV3	GCV4	NGCV2	NGCV3	NGCV4	OI
13x9	36	10	18.44	25.70	136.4	6.64	6.82	6.37	6.29
13x9	36	7.5			6.94			6.60	7.79
13x9	36	5	8.82	7.05	6.50	8.47	6.84	6.48	11.59
13x9	36	0	8.43	6.15	5.95	7.09	5.94	5.75	30.00
8x6	16	10	9.67	9.60	9.42	6.28	6.17	6.25	6.14
8x6	16	5	10.83	8.09	7.91	8.58	7.43	7.07	10.66
8x6	16	0	9.22	8.11	8.23	7.52	6.87	6.95	30.00
5x5	4	10	13.94	12.55	10.80	8.76	8.64	8.63	8.22
5x5	4	5	12.55	10.25	11.11	11.42	10.01	9.96	13.71
5x5	4	2.5	13.93	12.13	14.31	11.87	9.19	10.64	23.36

TABLE 2: rms errors in the corrected grid values obtained with various simulation runs. GCVm denotes GCV was used to estimate the smoothing parameter for the Laplacian smoothing spline of order m. NGCVm denotes GCV was not used with the Laplacian smoothing spline of order m. OI denotes the error estimate from Optimum Interpolation for the corresponding parameters. Other parameters used were $r_g = 30$, $r_o = 10$.

Type	Lambda ($A^b = A \times 10^b$)	rms error m = 2	rms error m = 3	rms error m = 4
Unified	GCV	13.94	12.55	10.80
$c_d = 10^{\circ}$	$25^{-6}, 25^{-5}, 25^{-4}$	8.76	8.64	8.63
Unified	GCV	12.55	10.25	11.11
$c_d = 5^{\circ}$	$25^{-6}, 25^{-5}, 25^{-4}$	11.42	10.01	9.96
Unified	GCV	13.93	12.13	14.31
$c_d = 2.5^{\circ}$	$25^{-6}, 1^{-3}, 1^{-3}$	11.87	9.19	10.64

Table 1C: rms errors of the analyzed values for GCV and non-GCV simulations on the 5x5 grid with 4 observation points. Specified error parameters were $r_g = 30$, $r_o = 10$.

Type	Lambda ($A^b = A\times10^b$)	rms error m = 2	rms error m = 3	rms error m = 4
No first-guess $c_d = 10^o$	GCV $5^{-3}, 5^{-2}, 2^{-1}$	8.65 8.55	7.63 7.44	8.42 7.63
Corrections to first-guess $c_d = 10^o$	GCV $3^{-2}, 1^{-1}, 1^{0}$	7.16 6.77	7.44 6.82	7.63 7.18
Unified $c_d = 10^o$	GCV $1^{-5}, 2^{-4}, 4^{-3}$	18.44 6.64	25.70 6.37	136.42 6.37
Unified $c_d = 7.5^o$	GCV 4^{-3}			6.94 6.60
Unified $c_d = 5^o$	GCV $25^{-6}, 25^{-5}, 25^{-4}$	8.82 8.47	7.05 6.84	6.50 6.48
Unified $c_d = 0^o$	GCV $25^{-6}, 25^{-5}, 25^{-4}$	8.43 7.09	6.15 5.94	5.95 5.75

Table 1A: rms errors of the analyzed values for GCV and non-GCV simulations on the 13x9 grid with 36 observation locations. Specified error parameters were $r_g = 30$, $r_o = 10$.

Type	Lambda ($A^b = A\times10^b$)	rms error m = 2	rms error m = 3	rms error m = 4
No first-guess $c_d = 10^o$	GCV $1^{-2}, 1^{-1}, 1^{-1}$	10.10 9.80	9.75 8.79	10.86 9.85
Corrections to first-guess $c_d = 10^o$	GCV $1^{-1}, 5^{-1}, 1^{0}$	7.00 6.48	7.58 6.60	10.38 8.25
Unified $c_d = 10^o$	GCV $2^{-5}, 3^{-4}, 8^{-3}$	9.67 6.28	9.60 6.17	9.42 6.25
Unified $c_d = 5^o$	GCV $5^{-5}, 4^{-4}, 25^{-4}$	10.83 8.58	8.09 7.43	7.91 7.07
Unified $c_d = 0^o$	GCV $4^{-5}, 4^{-4}, 4^{-3}$	9.22 7.52	8.11 6.87	8.23 6.95

Table 1B: rms errors of the analyzed values for GCV and non-GCV simulations on the 8x6 grid with 16 observation points. Specified error parameters were $r_g = 30$, $r_o = 10$.

Wahba G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross validation, Mon. Wea. Rev. 108, 1122-1143.

Wendelberger, J., 1981: The computation of Laplacian smoothing splines with examples, Tech. Rep. 648, Dept. Statistics, University of Wisconsin, Madison, 66 pp. [NTIS AD-A108 746/9].

Wendelberger, J., 1982: Smoothing noisy data with multidimensional splines and generalized cross-validation, Ph.D. thesis, Dept. Stat., University of Wisconsin, Madison, 336 pp. [No. 83-01893, University Microfilms, Ann Arbor, MI 48106].

REFERENCES

Barnes, S. L., 1973: Mesoscale objective map analysis using weighted time-series observations, NOAA Tech. Memo. ERL NSSL-62, 60 pp. [NTIS COM-73-10871].

Bergman, K. H. ,1979: Multivariate analysis of temperatures and wind using optimum interpolation, Mon. Wea. Rev. **107**, 1423-1444.

Cressman, G. P., 1959: An operative objective analysis scheme, Mon. Wea. Rev. **87**, 367-374.

Duchon, J., 1976: Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, R.A.I.R.O. Anal. Numer. **10**, 5-12.

Duchon, J., 1977: Splines minimizing rotation-invariant semi-norms in Sobolev spaces, in: W. Schempp and K. Zeller, eds., **Constructive Theory of Functions of Several Variables**, Lecture Notes in Math. **571**, Springer, 85-100.

Franke, R., 1985: Sources of error in objective analysis, Mon. Wea. Rev. **113**, 260-270.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Translated from Russian by Israel Program for Scientific Translations, 1965, 242 pp. [NTIS **TT65-50007**].

Harder, R. L. and R. N. Desmarais, 1972: Interpolation using surface splines, J. Aircraft **9**, 189-191.

Koehler, T. L., 1979: A case study of height and temperature analyses derived from Nimbus-6 satellite soundings on a fine mesh model grid, Ph.D. thesis, Dept. Meteor., University of Wisconsin, Madison, 186 pp., [No. **79-27181**, University Microfilms, Ann Arbor, MI 48106].

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme, Mon. Wea. Rev. **109**, 701-721.

Meinguet, J., 1979: Multivariate interpolation at arbitrary points made simple, Z. Agnew. Math. **30**, 292-304.

Meinguet, J., 1979a: An intrinsic approach to multivariate spline interpolation at arbitrary points, in: B. N. Sahney, ed., **Polynomial and Spline Approximation**, Reidel, 163-190.

Seaman, R. S., 1983: Objective analysis accuracies of statistical interpolation and successive correction schemes, Aust. Meteor. Mag. **31**, 225-240.

Thiebaux, H. J., 1985: On approximations to geopotential and wind-field correlation structures, Tellus **37a**, 126-131.

1983). This may be possible for LSS when a smoothing parameter is specified. While the simulation program is available and gives very good results (as can be seen in Figures 17-19 for OI), Seaman's approach requires considerably less computation.

All simulations reported on here were univariate. In its current practical applications OI is applied in a multivariate setting. As noted by Wahba and Wendelberger, LSS is also applicable in the multivariate setting, but the method has not been rigorously tested, since they computed only a small number of examples. There is no reason to suspect that LSS will perform any less well, compared to OI, in this setting than it does in the univariate case. It is necessary to perform some comparable analyses for the two methods to verify this, however, and such a study is anticipated in the near future.

5. Acknowledgements

Extensive use was made of the IBM 3033AP at the Naval Postgraduate School Computer Center. Dr. Edward Barker of the Naval Environmental Prediction Research Facility read preliminary drafts of the paper and his suggestions improved the final version of it.

general applicability and would be similar, independent of the source of the data. Nonetheless, the results will be discussed in terms of the setting in which they were performed. It is apparent that the routine, day-to-day use of GCV is not a suitable, nor cost effective way, to determine the smoothing parameter for LSS. On the other hand, it does seem to be useful to determine a single suitable value to use for all realizations in some particular ensemble. No effort was made to determine the optimum value of λ to use for any set of realizations in our simulation, although in some cases a set was run with more than one value of λ. The results indicated that the "eye-ball" average used was a good value, although it could be improved on if there is access to the actual errors. In practice, of course, this is not the case.

The use of LSS in a unified sense to treat both first-guess and observations in the same manner looks promising in regions where the observations are sparse. My investigation here is not really complete, however, and some additional work is necessary to verify the apparent conclusion that can be made. In particular, the simulations had perfect knowledge of the statistical characteristics of both the first-guess and observation error, and in practice this is impossible. An investigation of the sensitivity of both statistical interpolation and LSS to erroneous specification of the statistical characteristics of the errors is planned. In addition to this, several sets of grids with sparse observations will be used in the study. For statistical interpolation it is possible to find the rms errors over a given ensemble of realizations without simulation (see Seaman,

correlated first-guess errors (the more strongly correlated, the better), OI cannot be very successful in regions where the distance between observation points is a significant fraction of the correlation distance (or perhaps, where the density of observation points per unit of correlation distance area is small). This behavior is seen in the last column of Table 2, which also summarizes the results for m=4 on the 3 grids used in the simulations. Observe that no correction can be expected to be made if the first-guess errors are uncorrelated ($c_d = \emptyset$). The relationship is complex, as is seen through the inversion of the system of equaions for the coefficients in the approximation, and could be expected to depend heavily on distances to several nearby observation points as well as the first-guess grid size.

The phenomenon is more clearly illustrated by Figures 17-19, which graphically shows some of the data of Table 2. Figure 17 shows the rms errors as a function of correlation distance for unified LSS for m=4 with and without GCV, and for OI from simulations, along with the expected rms error from OI. Figures 18 and 19 show the corresponding data for the 8x6 grid and the 5x5 grid, again with m=4.

4. Conclusions

This investigation has been primarily concerned with the performance of generalized cross validation in conjunction with its use to determine the smoothing parameter for Laplacian smoothing splines applied to the objective analysis problem in numerical weather prediction. While the simulations performed have been within that context, I feel that the results have

parameter which might possibly be related to the minimum value of the function is the value of the smoothing parameter, however Figure 9 again shows no particular evidence of correlation. None of the available parameters seem to be indicative of "extreme" cases, and in particular are not detectable either from the GCV value or the smoothing parameter value. The one exception to that is the extreme point which is shown on the boundary, which does correspond to a very small value of the smoothing parameter, λ. This implies that little smoothing was applied for this particular realization. The case also corresponds to a relatively small value of the ratio of rms first-guess error to rms observation error.

Figures 11-16 show the corresponding plots for realizations incorporating uncorrelated first-guess error (correlation distance $c_d = 0$), again for the 13x9 grid. Except for there being no cases giving really poor performance of GCV in these realizations, the behavior is basically the same as Figures 5-10. The only evidence of correlated values is between the rms errors of the analyzed values with and without GCV. Other plots for variations in correlation distances, smoothness parameters, grids and observation point sets support these results.

The relative constancy of the the rms errors obtained by the LSS as the correlation distance is varied, as opposed to the rapid increase in errors obtained by OI as the correlation distance decreases is thought provoking. One is easily convinced that since OI is based on the idea of a correction to first-guess errors and since the successful application of OI depends on

variation of the rms errors in the analyzed values but the errors do not tend to increase greatly as correlation distance is decreased. The last result is discussed in more detail later.

In addition to the tabulation shown, a number of plots of various parameters versus rms error in the analyzed values for some of the sets of GCV realizations were made. Some of those are reproduced here, showing a typical range of behavior. In Figures 5-10, the simulations were on the 13x9 grid (Figure 1), with a correlation distance of $c_d = 7.5^{\circ}$, and smoothness parameter $m = 4$. One point is off the graph area and its projection onto the boundary is shown. Figure 5 shows the rms errors of the analyzed values for the non-GCV simulations versus the rms errors of the analyzed values for the corresponding GCV simulations. These appear to be correlated fairly well. The total rms error is smaller for analyses using a specified smoothing parameter value than for those obtained using GCV. Figures 5-10 show scatter diagrams of first-guess rms error, observation error, ratio of rms first-guess to rms observation error, log λ, and GCV function value, respectively, versus the rms error in the analyzed values obtained with GCV. No correlation between these sets of values is apparent, and in particular the GCV function value does not seem to be correlated with the actual rms errors in the analyzed values. Thus it would appear that while computing the GCV function gives one something to minimize, in this problem it is not true that the minimum of it corresponds to a minimum in the rms error of the analyzed values. This is further borne out by the generally smaller errors are obtained by specifying a constant value for the smoothing parameter. The other

Table 1 shows the results of LSS simulations with several different processes. The objective analysis processes used were (1) the Wahba and Wendelberger method with no first-guess and $c_d = 10^o$, (2) the correction to first-guess method with LSS applied to the first-guess error and $c_d = 10^o$, and (3) the unified scheme with various correlation distances specified. For the 13x9 and 8x6 grids, the three values, m=2, 3, and 4 were used. Wahba and Wendelberger had previously reported that for similar data, m=4 or 5 seemed to be appropriate. The simulations performed here indicate that for the particular underlying function used, m=3 or 4 is best. Though not discernable from the table, the GCV function generally was found to have multiple local minima, especially for the larger data sets when the first-guess errors were highly correlated (large c_d). The results for m=4 and $c_d = 10^o$ are not completely reliable since three of the cases failed in the determination of the smoothing parameter using GCV, and five others gave very poor results. The failures were probably caused by inexact computations of the square root of the correlation matrix in the LSS program, because the correlation matrix is poorly conditioned with respect to the precision used in the computations (double precision (REAL*8) on an IBM computer). No failures occurred when the smoothing parameter was specified.

The principal results to be drawn from Table 1 are: making corrections to the first-guess field always gave better analyzed values than not using the first-guess field, the unified scheme (without GCV) gave better analyzed values than the no first-guess process, and decreasing the correlation distance results in some

gressive correlation models approximate the actual first-guess error data better than Gaussian functions, and have other required properties needed in the multivariate case, as well. However, the overall results of this study would probably be altered only slightly by use of other correlation functions and distance in kilometers.

3. Results of the study

A number of simulations with different grid and observation point sets, correlation distances, and order of the LSS, were computed. A table giving the parameters of most of these simulations, including the resulting rms errors of the estimated grid point values is given in Table 1. All simulations used values of 30 m and 10 m for the standard deviations of the first-guess and observation errors, respectively.

The efficacy of the generalized cross validation (GCV) process as a scheme for choosing the smoothing parameter was one of the primary points investigated. An attempt was made to determine if the rms error resulting from the choice of smoothing parameter by GCV were related to any other parameters in the particular realization. With given parameters, a set of 50 (or 100, for the 3x6 and 5x5 grids) realizations were generated, and the rms errors of the analyzed values at the grid points, along with the rms first-guess and observation errors, the smoothing parameter value, and the GCV function value were tabulated. The realizations were then repeated using a smoothing parameter value determined from an eye-ball average of the λ values obtained through GCV over all realizations in the particular ensemble.

Before the system of equations can be solved for the coefficients the smoothing parameter λ must be specified. Wahba and Wendelberger (1980) show how to choose this smoothing parameter using GCV. In simple cross validation λ is selected to minimize the square of the errors in the scheme measured by sequentially predicting the value at each data point when it is omitted from the set, then summing over all data points. This turns out to be an unreasonably expensive calculation, and GCV is a procedure for estimating the minimizing parameter in the particular realization.

Since Optimum Interpolation (OI) is typically used in meteorological analysis, the performance of LSS and GCV was measured relative to that of OI. In the ensemble mean-squared error sense, OI must perform at least as well as any other scheme based on making corrections to a first-guess field. As was shown in Franke (1985), the simulation program yields rms errors which compare very favorably with the predicted values from the scheme, so it was not necessary to run the simulations for OI. The simulations for OI are quite inexpensive to compute, however, and some were run as a check of the simulation program. In all the simulations, the spatial covariance of the first-guess errors were assumed to be Gaussian,

$$C(P,Q) = r_g^2 \exp(-(||P-Q||/c_d)^2).$$

Here c_d is a parameter, referred to in the sequel as the correlation distance, and r_g^2 is the variance of the first-guess error. The use of Gaussian correlation functions and distance in degrees is not necessarily the best assumption that could be made. For example, recent work by Thiebaux (1985) has shown that autore-

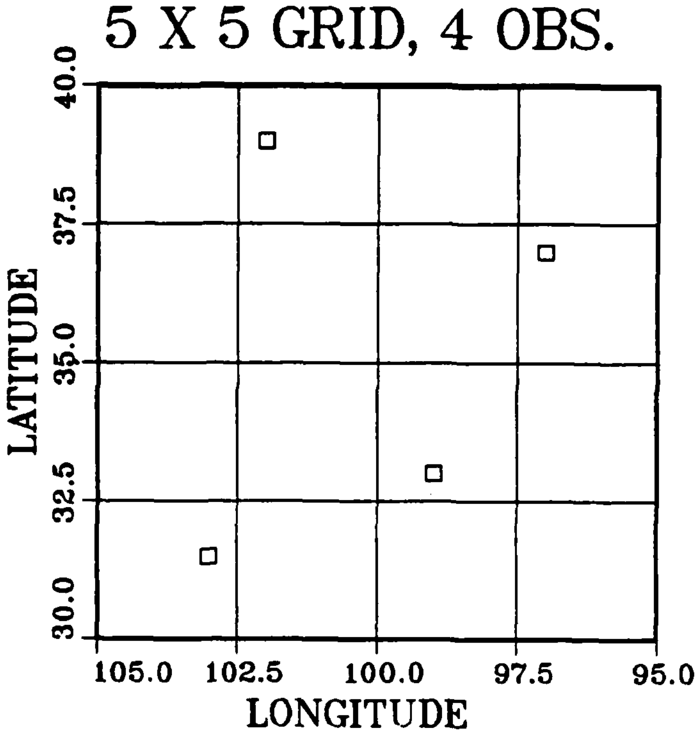


FIGURE 3

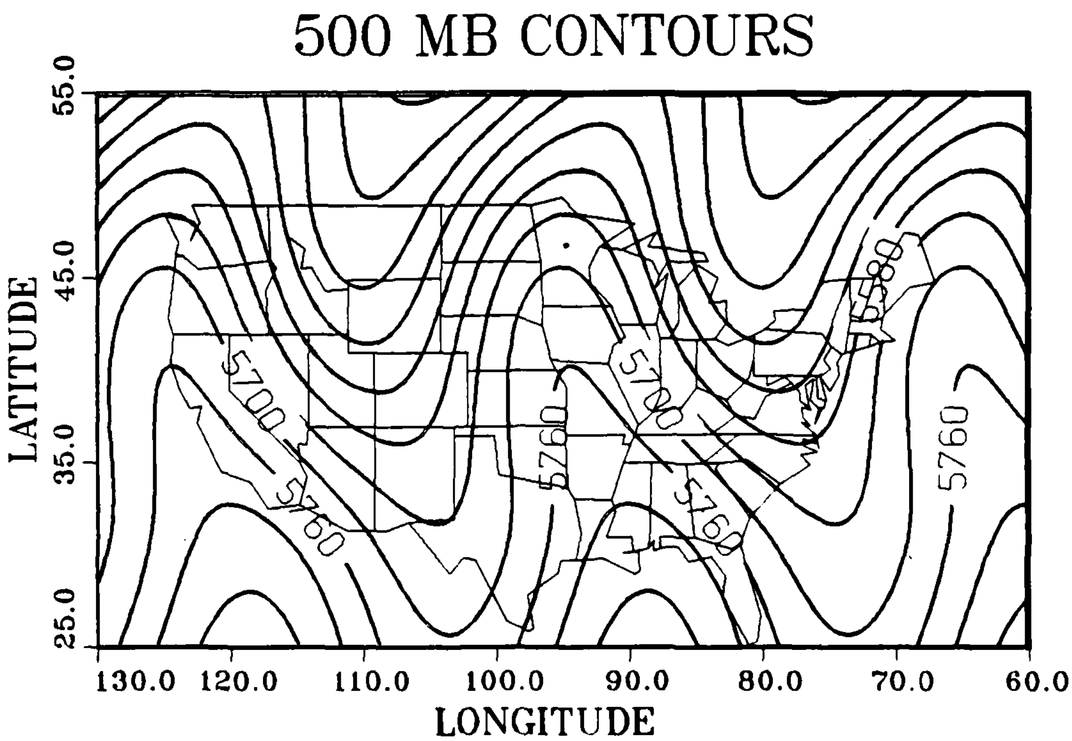


FIGURE 4

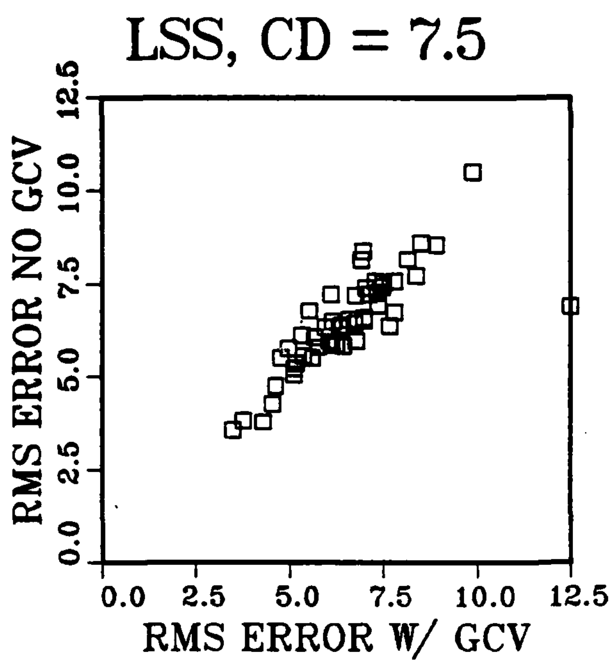


FIGURE 5

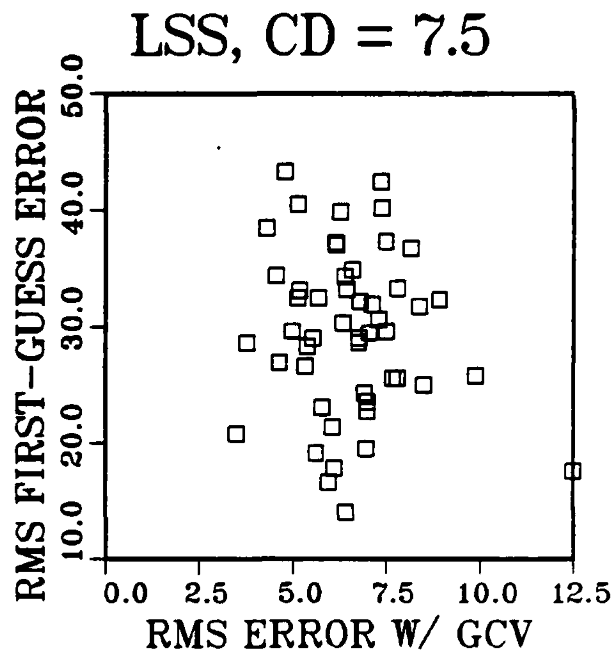


FIGURE 6

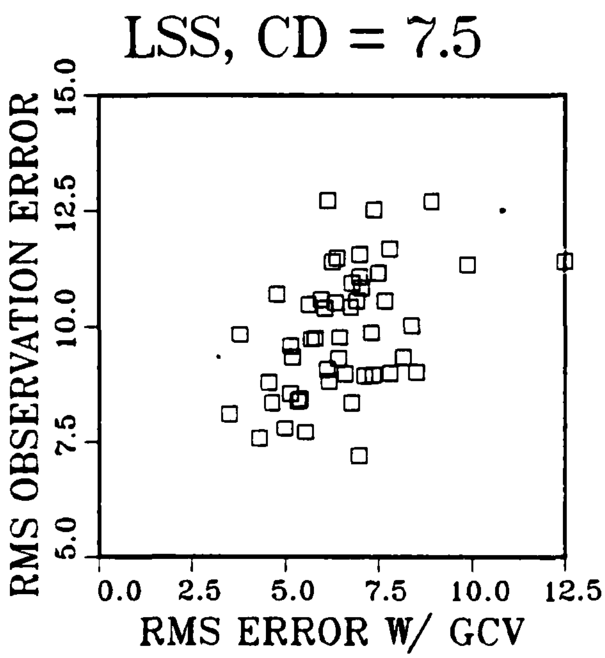


FIGURE 7

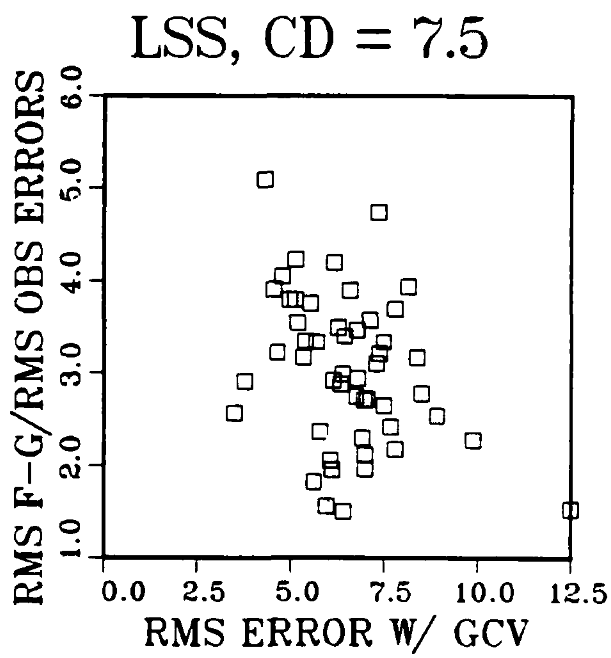


FIGURE 8

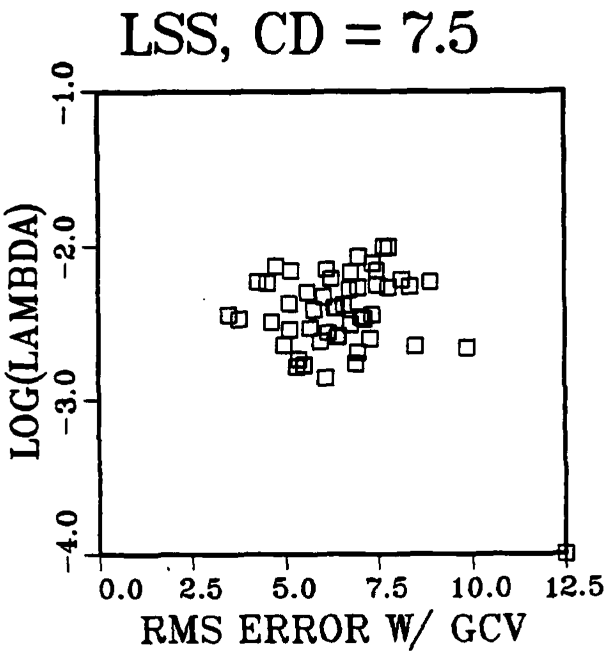


FIGURE 9

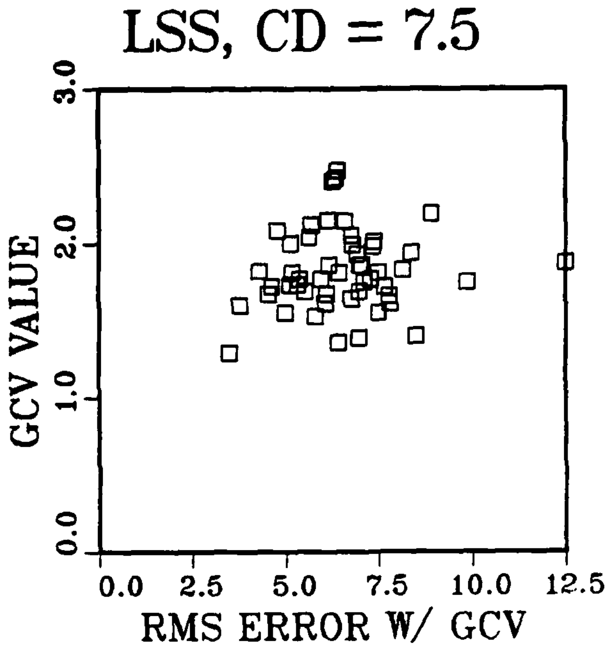


FIGURE 10

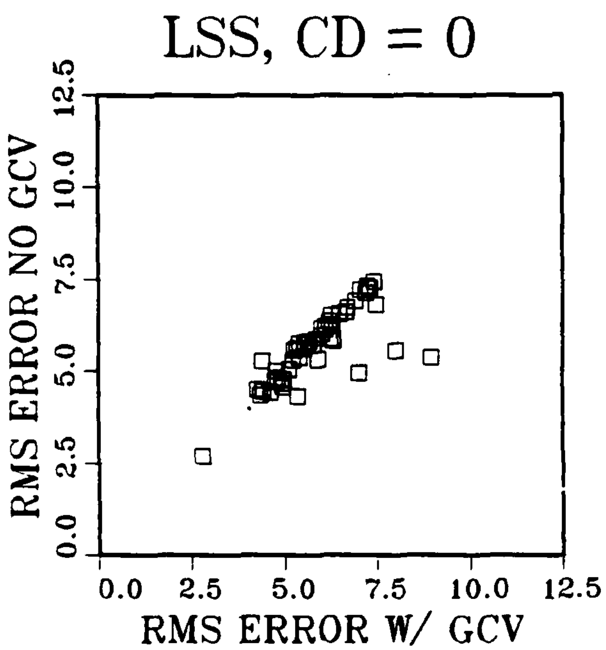


FIGURE 11

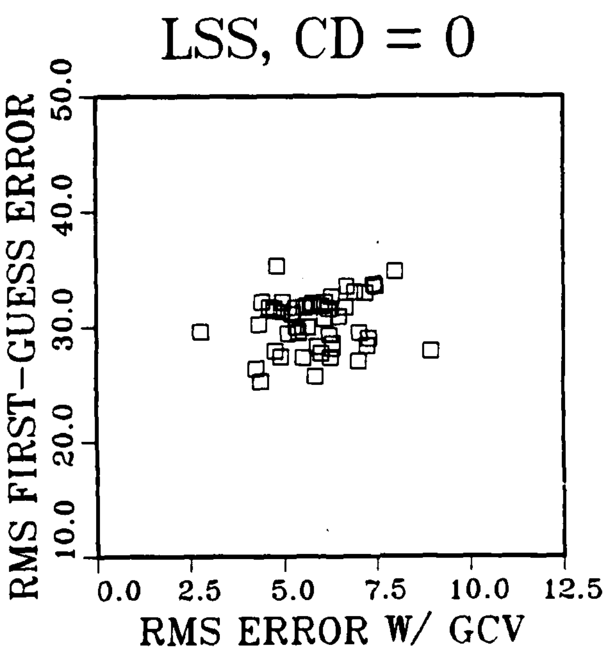


FIGURE 12

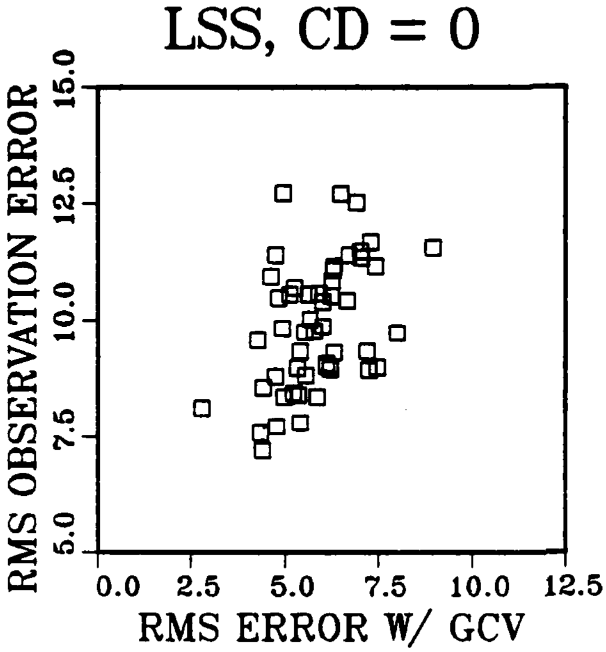


FIGURE 13

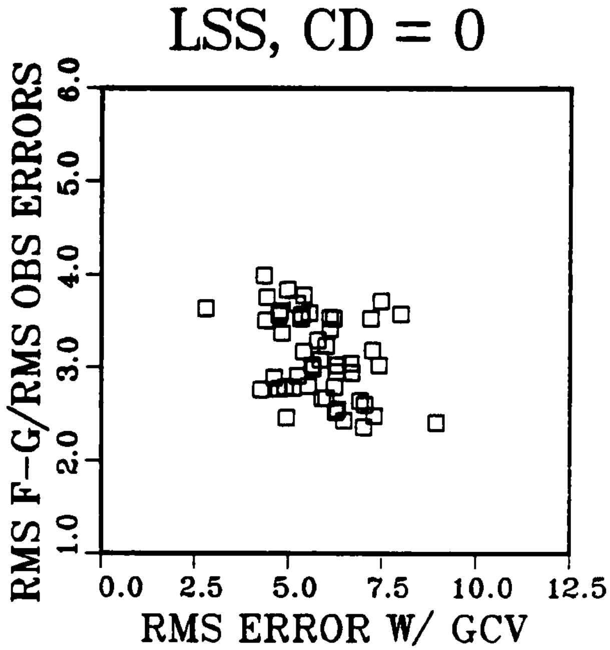


FIGURE 14

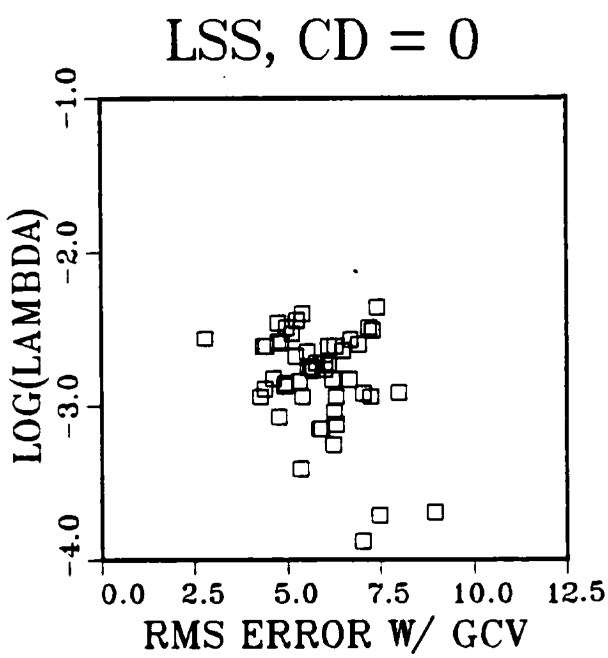


FIGURE 15

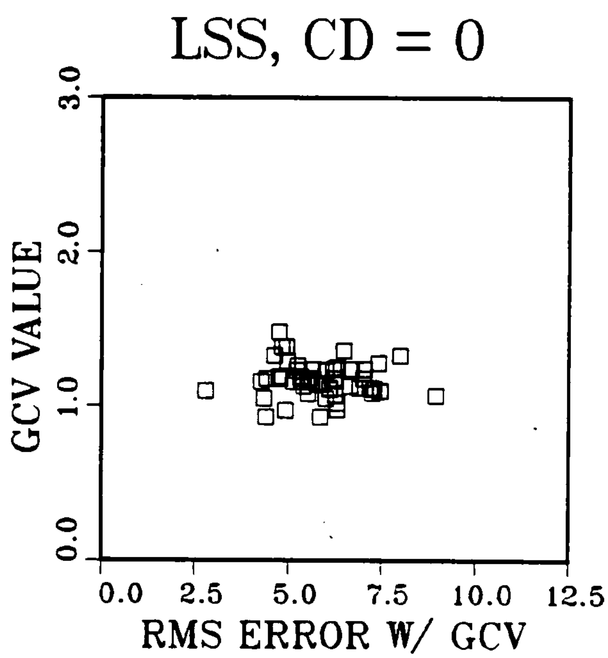


FIGURE 16

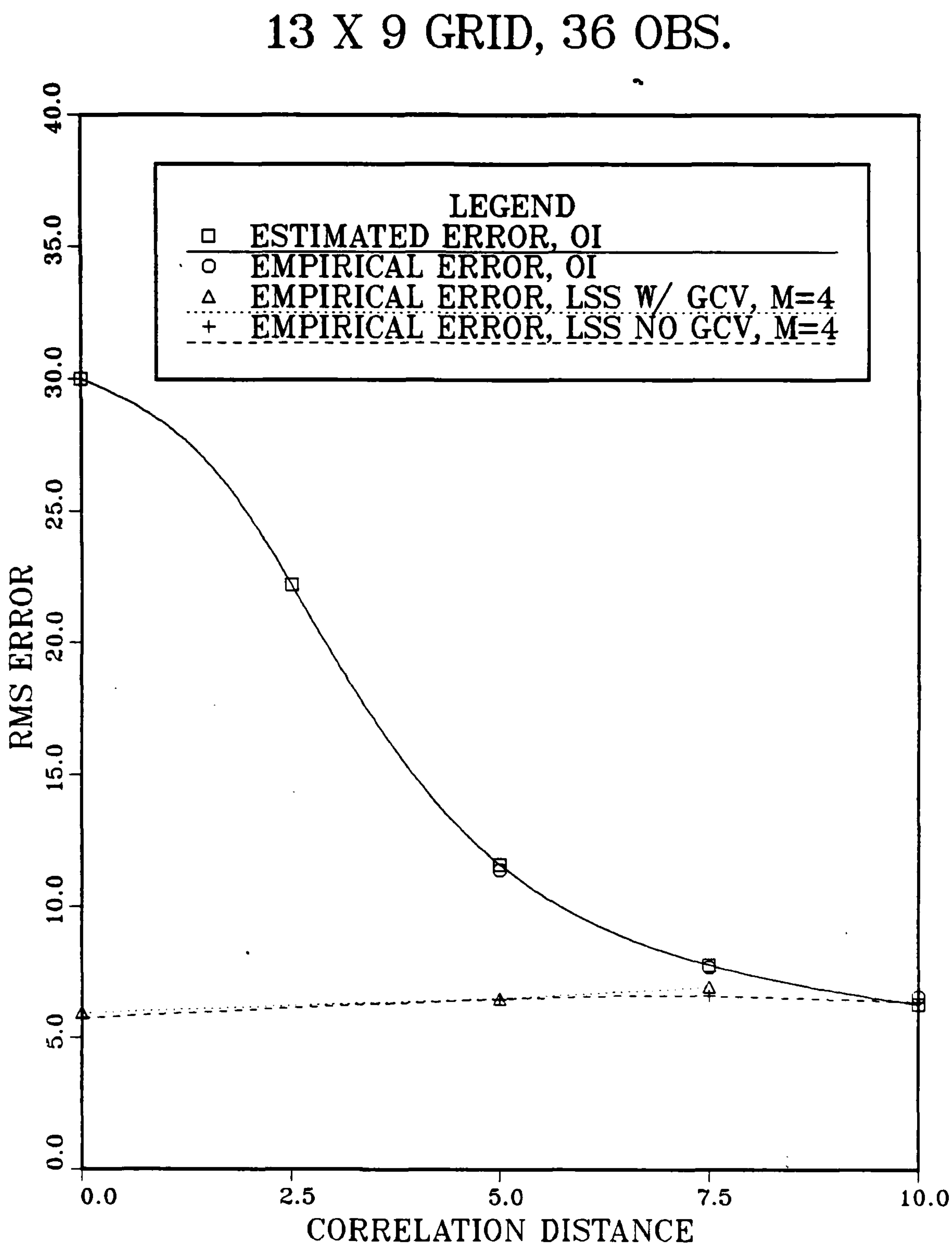

13 X 9 GRID, 36 OBS.
LEGEND
ESTIMATED ERROR, OI
EMPIRICAL ERROR, OI
EMPIRICAL ERROR, LSS W/ GCV, M=4
EMPIRICAL ERROR, LSS NO GCV, M=4
RMS ERROR
CORRELATION DISTANCE
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
2.5
7.5

FIGURE 17

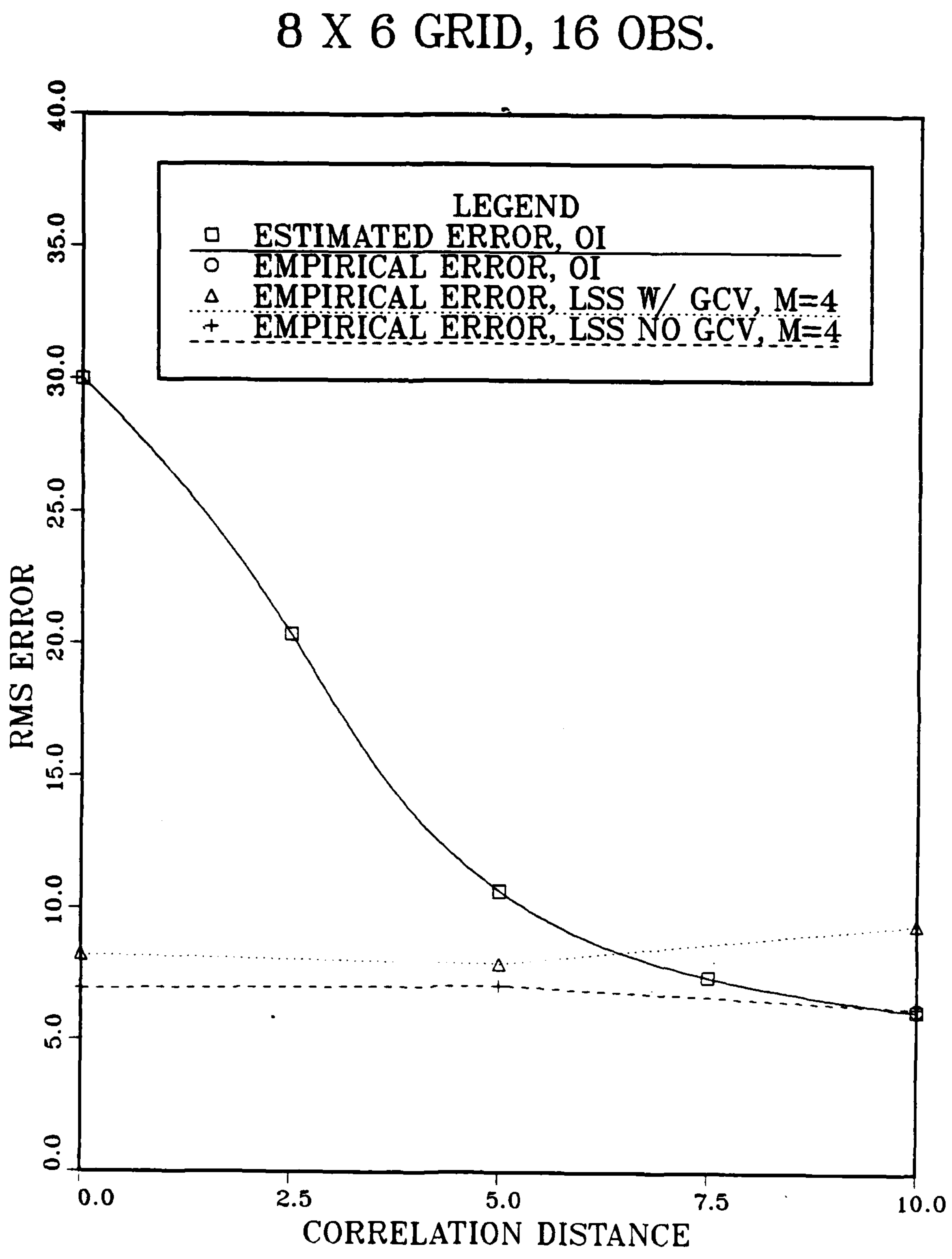
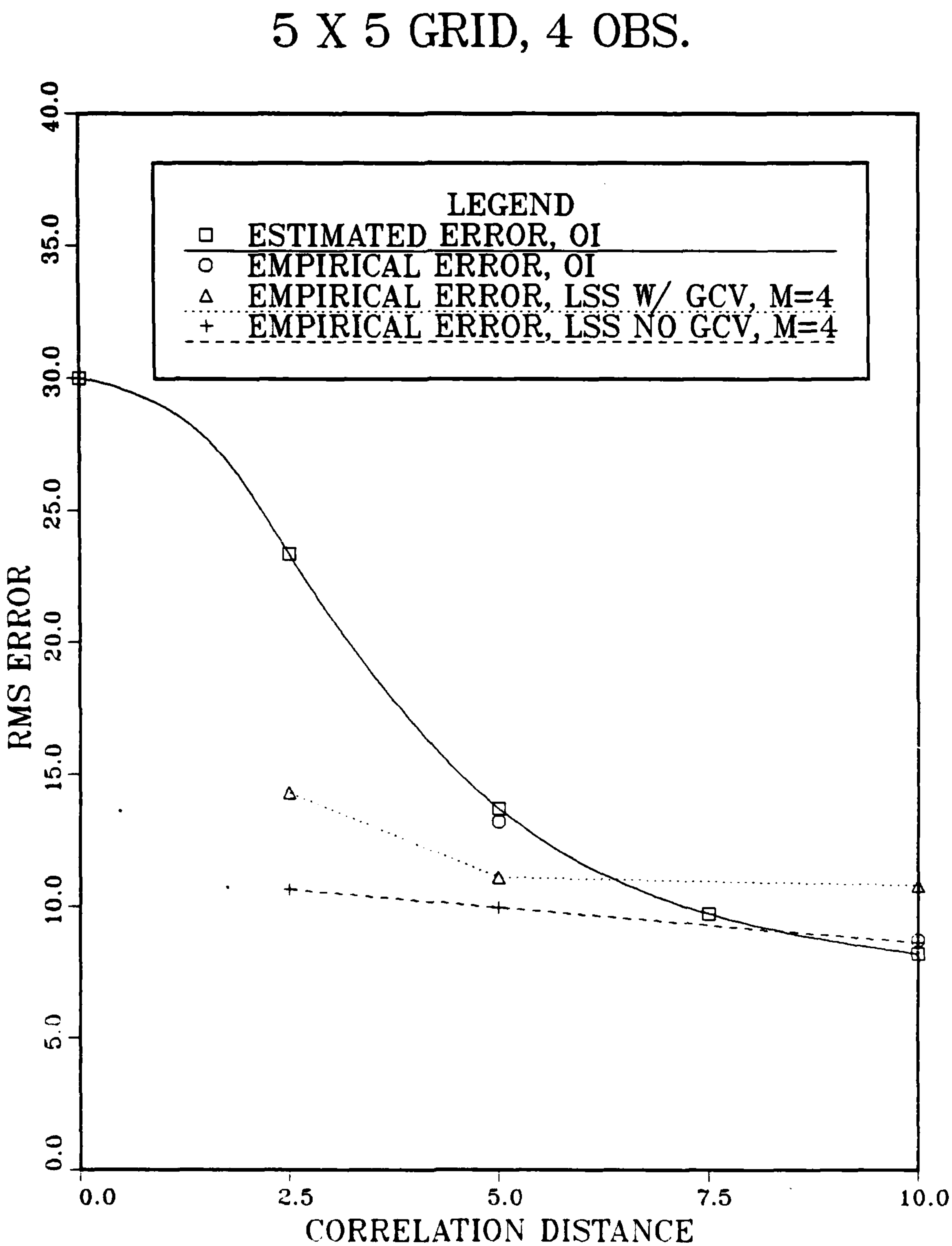

8 X 6 GRID, 16 OBS.
LEGEND
ESTIMATED ERROR, OI
EMPIRICAL ERROR, OI
EMPIRICAL ERROR, LSS W/ GCV, M=4
EMPIRICAL ERROR, LSS NO GCV, M=4
RMS ERROR
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
0.0
2.5
5.0
7.5
10.0
CORRELATION DISTANCE

FIGURE 18



5 X 5 GRID, 4 OBS.
LEGEND
ESTIMATED ERROR, OI
EMPIRICAL ERROR, OI
EMPIRICAL ERROR, LSS W/ GCV, M=4
EMPIRICAL ERROR, LSS NO GCV, M=4
RMS ERROR
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
0.0
2.5
5.0
7.5
10.0
CORRELATION DISTANCE

FIGURE 19

DISTRIBUTION LIST

ASST. FOR ENV. SCIENCES
ASST. SEC. OF THE NAVY (R&D)
ROOM 5E731, THE PENTAGON
WASHINGTON, DC 20350

CHIEF OF NAVAL RESEARCH (2)
LIBRARY SERVICES, CODE 784
BALLSTON TOWER #1
800 QUINCY ST.
ARLINGTON, VA 22217

OFFICE OF NAVAL RESEARCH
CODE 422AT
ARLINGTON, VA 22217

COMMANDING OFFICER
NORDA
NSTL, MS 39529

COMMANDING OFFICER
FLENUMOCEANCEN
MONTEREY, CA 93940

SUPERINTENDENT
LIBRARY REPORTS
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

NAVAL POSTGRADUATE SCHOOL
METEOROLOGY DEPT.
MONTEREY, CA 93943

NAVAL POSTGRADUATE SCHOOL
OCEANOGRAPHY DEPT.
MONTEREY, CA 93943

LIBRARY (2)
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

COMMANDER (2)
NAVAIRSYSCOM
ATTN: LIBRARY (AIR-7226)
WASHINGTON, DC 20361

COMMANDER
NAVAIRSYSCOM (AIR-330)
WASHINGTON, DC 20361

COMMANDER
NAVOCEANSYSCEN
DR. J. RICHTER, CODE 532
SAN DIEGO, CA 92152

USAFETAC/TS
SCOTT AFB, IL 62225

SUPERINTENDENT
ATTN: USAFA (DEG)
USAF ACADEMY, CO 80840

AFGWC/DAPL
OFFUTT AFB, NE 68113

AFGL/LY
HANSCOM AFB, MA 01731

AFGL/OPI
HANSCOM AFB, MA 01731

OFFICER IN CHARGE
SERVICE SCHOOL COMMAND
DET. CHANUTE/STOP 62
CHANUTE AFB, IL 61868

COMMANDER & DIRECTOR
ATTN: DELAS-D
U.S. ARMY ATMOS. SCI. LAB
WHITE SAND MISSILE RANGE
WHITE SANDS, NM 88002

DIRECTOR (2)
DEFENSE TECH. INFORMATION
CENTER, CAMERON STATION
ALEXANDRIA, VA 22314

NOAA-NESDIS LIAISON
ATTN: CODE SC2
NASA-JOHNSON SPACE CENTER
HOUSTON, TX 77058

DIRECTOR
NATIONAL METEORO. CENTER
NWS, NOAA
WWB W32, RM 204
WASHINGTON, DC 20233

ACQUISITIONS SECT. IRDB-D823
LIBRARY & INFO. SERV., NOAA
6009 EXECUTIVE BLVD.
ROCKVILLE, MD 20852

DIRECTOR
OFFICE OF PROGRAMS RX3
NOAA RESEARCH LAB
BOULDER, CO 80302

NATIONAL WEATHER SERVICE
WORLD WEATHER BLDG., RM 307
5200 AUTH ROAD
CAMP SPRINGS, MD 20023

DIRECTOR
NATIONAL SEVERE STORMS LAB
1313 HALLEY CIRCLE
NORMAN, OK 73069

DIRECTOR
GEOPHYS. FLUID DYNAMICS LAB
NOAA, PRINCETON UNIVERSITY
P.O. BOX 308
PRINCETON, NJ 08540

DIRECTOR
TECHNIQUES DEVELOPMENT LAB
GRAMAX BLDG.
8060 13TH ST.
SILVER SPRING, MD 20910

LABORATORY FOR ATMOS. SCI.
NASA GODDARD SPACE FLIGHT CEN.
GREENBELT, MD 20771

EXECUTIVE SECRETARY, CAO
SUBCOMMITTEE ON ATMOS. SCI.
NATIONAL SCIENCE FOUNDATION
RM. 510, 1800 G. STREET, NW
WASHINGTON, DC 20550

COLORADO STATE UNIVERSITY
ATMOSPHERIC SCIENCES DEPT.
ATTN: DR. WILLIAM GRAY
FORT COLLINS, CO 80523

ATMOSPHERIC SCIENCES DEPT.
UCLA
405 HILGARD AVE.
LOS ANGELES, CA 90024

CHAIRMAN, METEOROLOGY DEPT.
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73069

CHAIRMAN, METEOROLOGY DEPT.
CALIFORNIA STATE UNIVERSITY
SAN JOSE, CA 95192

COLORADO STATE UNIVERSITY
ATMOSPHERIC SCIENCES DEPT.
ATTN: LIBRARIAN
FT. COLLINS, CO 80523

NATIONAL CENTER FOR ATMOS.
RSCH., LIBRARY ACQUISITIONS
P.O. BOX 3000
BOULDER, CO 80302

UNIVERSITY OF WASHINGTON
ATMOSPHERIC SCIENCES DEPT.
SEATTLE, WA 98195

CHAIRMAN, METEOROLOGY DEPT.
PENNSYLVANIA STATE UNIV.
503 DEIKE BLDG.
UNIVERSITY PARK, PA 16802

FLORIDA STATE UNIVERSITY
ENVIRONMENTAL SCIENCES DEPT.
TALLAHASSEE, FL 32306

UNIVERSITY OF HAWAII
METEOROLOGY DEPT.
2525 CORREA ROAD
HONOLULU, HI 96822

DIRECTOR
COASTAL STUDIES INSTITUTE
LOUISIANA STATE UNIVERSITY
ATTN: O. HUH
BATON ROUGE, LA 70803

ATMOSPHERIC SCIENCES DEPT.
OREGON STATE UNIVERSITY
CORVALLIS, OR 97331

UNIVERSITY OF MARYLAND
METEOROLOGY DEPT.
COLLEGE PARK, MD 20742

CHAIRMAN
ATMOS. SCIENCES DEPT.
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VA 22903

CHAIRMAN
METEOROLOGY DEPT.
MASSACHUSETTS INSTITUTE OF
TECHNOLOGY
CAMBRIDGE, MA 02139

CHAIRMAN, METEOROLOGY DEPT.
UNIVERSITY OF UTAH
SALT LAKE CITY, UT 84112

ATMOSPHERIC SCIENCES CENTER
DESERT RESEARCH INSTITUTE
P.O. BOX 60220
RENO, NV 89506

ATMOSPHERIC SCI. RSCH. CENTER
NEW YORK STATE UNIV.
1400 WASHINGTON AVE.
ALBANY, NY 12222

SYSTEMS & APPLIED SCI. CORP.
ATTN: LIBRARY, SUITE 500
6811 KENILWORTH AVE.
RIVERDALE, MD 20840

CENTER FOR ENV. & MAN, INC.
RESEARCH LIBRARY
275 WINDSOR ST.
HARTFORD, CT 06120

METEOROLOGY RESEARCH, INC.
464 W. WOODBURY RD.
ALTADENA, CA 91001

METEOROLOGY INTL., INC.
P.O. BOX 22920
CARMEL, CA 93922

CONTROL DATA CORP.
METEOROLOGY DEPT. RSCH. DIV.
2800 E. OLD SHAKOPEE RD.
BOX 1249
MINNEAPOLIS, MN 55440

SCIENCE APPLICATIONS, INC.
205 MONTECITO AVENUE
MONTEREY, CA 93940

OCEAN DATA SYSTEMS, INC.
2460 GARDEN ROAD
MONTEREY, CA 93940

EUROPEAN CENTRE FOR MEDIUM
RANGE WEATHER FORECASTS
SHINFIELD PARK, READING
BERKSHIRE RG29AX, ENGLAND

DEAN OF RESEARCH (2)
CODE 012
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

DEPT. OF MATHEMATICS
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

PROFESSOR G. LATTA, CODE 53LZ
CHRMN, DEPARTMENT OF MATH
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

PROFESSOR R. FRANKE, CODE 53FE
DEPARTMENT OF MATHEMATICS
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

DR. RICHARD LAU
OFFICE OF NAVAL RESEARCH
800 QUINCY ST.
ARLINGTON, VA 22217

PROFESSOR R.E. BARNHILL
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UT 84112

PROFESSOR G.M. NIELSON
DEPARTMENT OF MATHEMATICS
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85281

LIBRARY, FLEET NUMERICAL(2)
OCEANOGRAPHY CENTER
MONTEREY, CA 93943

DR. EDWARD BARKER
NAVAL ENVIRONMENTAL PREDICTION
RESEARCH FACILITY
MONTEREY, CA 93943

DR. THOMAS ROSMOND
NAVAL ENVIRONMENTAL PREDICTION
RESEARCH FACILITY
MONTEREY, CA 93943

PROFESSOR W.J. GORDON
CENTER FOR SCIENTIFIC
COMPUTATION & INTERACTIVE
GRAPHICS
DREXEL UNIVERSITY
PHILADELPHIA, PA 19104

MS. LINDA THIEL
DEPARTMENT OF MATHEMATICAL
SCIENCES
DREXEL UNIVERSITY
PHILADELPHIA, PA 19104

PROFESSOR GRACE WAHBA
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WI 53705

PROFESSOR DONALD GAVER
DEPARTMENT OF OPERATIONS
RESEARCH
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

PROFESSOR P.A. JACOBS
DEPARTMENT OF OPERATIONS
RESEARCH
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

PROFESSOR H. JEAN THIEBAUX
DEPARTMENT OF MATHEMATICS,
STATISTICS & COMPUTING SCIENCE
DALHAUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA,
CANADA 133H 4H8

R. R.S. SEAMAN
USTRALIAN NUMERICAL
METEOROLOGY RESEARCH CENTRE
.O. BOX 5089AA
ELBOURNE, VICTORIA,
USTRALIA, 3001

PROFESSOR PETER ALFELD
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UT 84112

PROFESSOR MARK E. HAWLEY
DEPARTMENT OF ENVIRONMENTAL
SCIENCES
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VA 22903

R. LAWRENCE BREAKER
EPARTMENT OF OCEANOGRAPHY
AVAL POSTGRADUATE SCHOOL
ONTEREY, CA 93943

PROFESSOR JAMES J. O'BRIEN
DEPARTMENT OF METEOROLOGY
THE FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306

MR. ROSS HOFFMAN
ATMOSPHERIC AND ENVIRONMENTAL
RESEARCH, INC.
840 MEMORIAL DRIVE
CAMBRIDGE, MA 02139

R. JAMES GOERSS
AVAL ENVIRONMENTAL PREDICTION
RESEARCH FACILITY
ONTEREY, CA 93943

1STLT KEN GALLUPPI
USAFETAC/DNO
SCOTT AFB, IL 62225-5438

DR. LICIA LENARDUZZI
IAMI
VIA CICOGNARA 7
20129 MILANO
ITALIA

R. PAUL F. TWITCHELL
AVAIRSYSCOM
IR-330G
ASHINGTON, DC 20361

RESEARCH ADMINISTRATION
CODE 012CW
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943

REPRODUCED AT GOVERNMENT EXPENSE

END

FILMED

10-85

DTIC